IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 16 December 2023, accepted 9 January 2024, date of publication 17 January 2024, date of current version 26 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3355194

RESEARCH ARTICLE

Convex Optimization Based High-Order Fuzzy Cognitive Map Modeling and Its Application in Time Series Predicting

DAN SHAN[1], LI WANG[1], WEI LU[2], (Member, IEEE), AND JUN CHEN[3]
[1]Department of Electronic Engineering, Dalian Neusoft University of Information, Dalian 116023, China
[2]School of Control Science and Engineering, Dalian University of Technology, Ganjingzi, Dalian, Liaoning 116024, China
[3]School of Materials, Dalian University of Technology, Ganjingzi, Dalian, Liaoning 116024, China

Corresponding author: Jun Chen (chenjun@dlut.edu.cn)

This work was supported by the Science and Technology Major Project of Xinjiang Uygur Autonomous Region under Grant 2022A01001.

ABSTRACT As a soft computing method, applying fuzzy cognitive map (FCM) to time series prediction has become a timely issue pursued by numerous researchers. Although many FCM construction methods have emerged, most of them exhibit obvious limitations in weight learning especially for long-term or complex time series. Either the weight calculation is computationally expensive, or it cannot achieve gratifying accuracy. In this paper, a new method for constructing FCM is proposed which extracts concepts from data by exploiting triangular membership function, and the weights of high-order FCM are subtly obtained by transforming the learning problem of FCM into a convex optimization problem with constraints. Since then, FCM with optimized weights is used to represent fuzzy logical relationships of time series and implement prediction further. Fifteen benchmark time series,such as Soybean Price time series, Yahoo stock time series, Condition monitoring of hydraulic systems time series etc. are applied to verify prediction performance of the proposed method. Accordingly, experiment results show that the proposed numerical prediction method of time series is effective and can acquire better prediction accuracy with lower computation time than other recent advanced methods. In addition, the influence of parameters of the method is analyzed individually.

INDEX TERMS Fuzzy cognitive map, convex optimization, time series.

I. INTRODUCTION

Fuzzy cognitive map(FCM) as an effective one of many soft computing approaches, such as genetics algorithm, neural network, fuzzy logic, fuzzy neural network and evolutionary algorithm, etc. [1], is a combination of fuzzy logic and network for knowledge representation and reasoning of causality. Firstly put forward by Kosko [2] as an extension of Axelord cognitive maps [3], FCM is a fuzzy weighted directed graph with nodes representing concepts and edge reflecting causal relationships between nodes, which has become a hot topic in the field of time series based modeling [4], [5], prediction [6], [7], decision-making [8], [9], and bioinformatics [55] for researchers. In predictive applications, some methods based on fuzzy time series have been presented and appreciated in many areas [36], [37], [38], [44], [45], such as stock price, university enrollments, economic growth, etc. Due to the excellent ability of FCM to capture the dynamic characteristics of a given system and achieve knowledge reasoning [39], recently most researchers associated with time series prediction are committed to the study of FCM weight learning methods. Papageorgiou reviewed some existing learning algorithms for FCMs [10] and Orang presents an up-to-date and comprehensive presentation of the theory and applications of FCMs [40], [41]. Among them, some algorithms laid on population perform well on reappearing target sequences, such as Stach et al. [11], RCGA-based divide-and-conquer method [12], Liu et al. [13], Zou and Liu [14], etc. Further, some hybrid algorithms are presented combined the idea of iteration and population, Such as Papageorgiou and Groumpos [15], Zhu and Zhang [16], Liu and Zhang [17], [32]. However, these algorithms all have a latent danger, which is that they may fall into local optimal solution. The

The associate editor coordinating the review of this manuscript and approving it for publication was Alba Amato.

reason for this is that the population algorithms utilize both group and individual optimal information during the iteration process, and its particle diversity rapidly disappears near local extremum, which cannot guarantee convergence to the global optimum, leading to potential limitations in model accuracy. In recent years, some new algorithms are proposed, such as quantum fuzzy cognitive maps(QFCM) [46], interval type-2 fuzzy cognitive maps(IT2FCM) [54] and neuro-fuzzy cognitive maps(NFCM) [47]. QFCM is presented to provide the facility for simultaneous static and dynamic analyses. IT2FCM applied in cognitive controller provides a bridge between the well-developed cognitive sciences and control theory. NFCM is mostly suited to model complex nonlinear technical systems with dynamic internal characteristics. Although the effectiveness of modeling and forecasting is constantly improving with the continuous emergence of these new algorithms, QFCM, IT2FCM and NFCM still have common problems, which are high computational and time-consuming, especially in the case that the number of FCM nodes increases and the corresponding number of weights increases dramatically.

In order to reduce computational complexity, least square based LSFCM [36] and partitioning strategies based FCM [42] are proposed. Similar to most methods, FCM nodes are obtained by clustering time series [36]. In the LSFCM modeling process, first, fuzzy c-means clustering is employed to obtain concepts, and then the least squares method is applied to calculate weights, which is a direct and one-time solution of matrix equation without repetitious stochastic searching. However, to obtain the optimal concepts, particle swarm optimization is required to improve the prediction accuracy. Similarly, in the strategies based FCM modeling process, fuzzy c-means clustering is employed. Different from LSFCM, fuzzy c-means clustering here is used to divide the time series into multiple subsequences, and FCMs are constructed in terms of these subsequences respectively. Finally, the FCM models are merged by fuzzy rules.

Moreover, learning FCM weights require historical time series data, which refers to fuzzy time series transformed from classical ones. Compared with classical ones, fuzzy time series possess the characteristic of linguistic values rather than numeric values, and are capable of handling ambiguities. Fuzzy time series also inherit the merit of classical ones, which can also be understood as being collected at fixed time intervals and independent of time. Therefore, fuzzy time series have some inherent, implicated and meaningful essence for prediction. Originally, fuzzy time series are proposed by Song and Chissom [18], [19], [20] engaging the formalisms of fuzzy set theory [21], which has been widely applied and appreciated in fields of enrollments [22], [23], [24], stock index [23], [24], electricity load [23], [24] etc. For time series prediction, Song and chissom divide it into four steps: (1) define the domain U and divide it into several intervals; (2) fuzzify the classical time series; (3) establish the fuzzy relation; (4)predict and defuzzify. Except for (3), (2) and (4) have a significant impact on the prediction accuracy as well. At present, the most widely used fuzzy and defuzzy method is c-means clustering algorithm [18], [25], [26], [27], [28], [29]. By optimizing the objective function, the memberships of each sample to all cluster centers are obtained. As an unsupervised method of deriving membership, c-means clustering has good flexibility [32]. However it also has an obvious disadvantage, that is, when it is used to fuzzify historical data, deviation is introduced. The deviation is also reflected in the process of defuzzification. As a result, deviation from historical data to fuzzy time series leads to inaccurate historical data restored. To some extent, the accuracy of prediction is degraded.

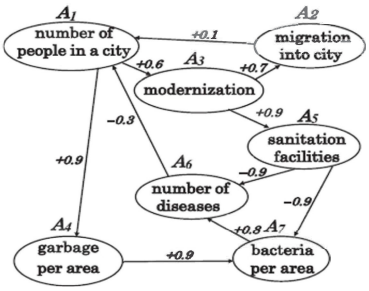
The aforementioned methods have achieved good results in time series prediction, yet they have three conspicuous problems. The first one is that during reasoning, the value of each FCM node at moment t is only related to the value at moment $t-1$, which is unreasonable. In reality, the value at moment t is related not only to moment $t-1$ but also to earlier moments. The second one is that during the weight learning processes, most of the methods involve population algorithms, potential local optimization problems may have adverse effects on FCM prediction performance. The third one is that algorithms based on neural networks, are generally time-consuming and labor-intensive. Therefore, In order to solve the problems found in the aforementioned methods, i.e. insufficient consideration of historical information, limited reasoning ability and performance accuracy, as well as large computational load and long computation time, we extend the 1-order FCM to high-order FCM by modifying the iteration formula to further enhance the representation ability of FCM. Further, we introduce convex optimization to solve the weights of the high-order FCM, which not only enhances the reasoning capacity and modeling precision by obtaining global optimal weight solution, but also reduces computational complexity.

In comparison with aforementioned methods of learning FCMs, the proposed method exhibits several appealing advantages.

1) A high order FCM was proposed, which not only considers the influence of the current moment on the concept values, but also takes into account the influence of previous moments on the concept values. It significantly improves the representation and inference ability of the FCM, and promotes the accuracy of FCM based time series prediction methods.

2) The proposed high order FCM weight learning algorithm is fast and effective because it transforms the weight learning problem of FCM into a constrained convex optimization problem, which can be solved by applying gradient methods and has polynomial time complexity.

3) The proposed method has a certain degree of interpretability. It converts time series into semantic time series by using a series of triangular fuzzy sets with semantic descriptions, and uses FCM to achieve semantic level inference and prediction. Therefore, the proposed method has a certain degree of interpretability.



(a) Fuzzy cognitive map model

	A_1	A_2	A_3	A_4	A_5	A_6	A_7
A_1	0	0.1	0	0	0	−0.3	0
A_2	0	0	0.7	0	0	0	0
A_3	0.6	0	0	0	0	0	0
A_4	0.9	0	0	0	0	0	0
A_5	0	0	0.9	0	0	0	0
A_6	0	0	0	0	−0.9	0	0.8
A_7	0	0	0	0.9	−0.9	0	0

(b) Relationship matrix

FIGURE 1. Fuzzy cognitive map model and its relationship matrix for public city health issues.

The remainder of this paper is organized as follows: In Section II, the preliminaries of FCMs from 1-order to high order are introduced. Subsequently, the detailed proposed method is provided in Section III. In section IV, fifteen benchmark time series data sets are used to validate feasibility and effectiveness of the proposed method. Finally, Section V concludes this paper.

II. FCMs FROM 1-ORDER TO HIGH ORDER

FCM is an effective and efficient tool for human knowledge representing and reasoning. Furthermore, FCM is very suitable for representing the fuzzy causal relationship between concepts, and the relationship between concepts has the internal characteristics of universality. Meanwhile, the more fuzzy the representation of knowledge, the easier it is to get knowledge from it. FCM model can simulate and analyze the behavior of real system, including periodic oscillation and other complex phenomena. Moreover, it can realize the matrix reasoning and make the reasoning process more concise.

FCM spreads out its appearance by graphical diagram with fuzzy causal concepts and mutual causal relationships among them. The interactions between concepts are directional, and the degrees of interactions are expressed by weights. FIGURE.1(a) shows an example of 1-order FCM model for public city health issues [30]. Nodes of FCM stand for concepts of the mimic dynamic system, such as variables, events, goals, etc. to be investigated. Simultaneously, edges of FCM show the strength of interaction between nodes in three type weights, namely positive, negative, and neutral. Quantitatively activation values of nodes are positioned in [0, 1], and the numeric values of weights are quantified into the interval of $[-1, +1]$. Symbolically, values of nodes are denoted by $A_i (i = 0, 1, \ldots, n)$, where n is the number of nodes. In FIGURE.1, n is 7. Similarly, values of weights are denoted by w_{ij}, which means the direction and strength of connection from node A_j to A_i. When $w_{ij} > 0$ like from A_4 to A_7, an increase of A_4 leads to an increase of A_7 by extent 0.9, oppositely a decrease of A_4 leads to a decrease of A_7. Similarly, when $w_{ij} < 0$ like from A_6 to A_1, an increase of A_6 leads to a decrease of A_1 by extent 0.3, oppositely a decrease of A_6 leads to an increase of A_1. Particularly when $w_{ij} = 0$, it corresponds to the removed edges between the nodes, which means no relationship from the start node to the target one. In some special cases, the start node is the same as the target node. In addition, the fuzzy causality between concepts in FCM can be expressed not only by directed graph, but also by weighted adjacency matrix referred to relationship matrix often, as shown in FIGURE.1(b).

Alternatively, the dynamics of FCM can be described in mathematical form as follows:

$$\mu_i(t+1) = f\left(\sum_{j=1}^{n} (w_{ij}\mu_j(t))\right) \tag{1}$$

where $\mu_i(t+1)$ is the activation level of ith node at moment $t+1$, and $\mu_j(t)$ is the activation level of jth node at moment t, $w_{ij} \in [-1, 1]$ is the weight from jth node to ith node, i.e., causality from A_j to A_i, n is the number of nodes and f is a nonlinear continuous non-decreasing transformation function, which has several options, such as bivalent function, trivalent function and sigmoid function, etc. Since both the sigmoid function itself and its inverse function are monotonically increasing, it is commonly chosen to be the activation function in FCM modeling and predicting. The expression of sigmoid function is shown in Eq.2.

$$f(x) = \frac{1}{1 + e^{-\lambda x}} \tag{2}$$

where λ is steepness parameter to provide some additional augmentation of the concept value. The larger the value of this parameter, the steeper the shape of the activation function, and the more sensitive it is to the value of x.

In order to improve the approximation ability of FCM to describe dynamic system, for $t+1$ moment, the value of each concept not only depends on the values of concepts at the immediate previous moment t, but also has a certain relationship with the values before t moment. The closer to the current moment is, the greater the impact is. In view of this consideration, it is necessary to increase the order of FCM. Assuming the order of FCM is q, there are q corresponding weights, namely the weight matrix w_{1ij} at the current time t, the weight matrix w_{2ij} at the previous time $t-1$, and so on, until the weight matrix w_{qij} at time $t-q+1$. Here, for all w_{1ij},w_{2ij},…,w_{qij} are within the interval [−1,1], where $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, n$.

$$\mu_i(t+1) = f\left(\sum_{j=1}^{n} (w_{1ij}\mu_j(t)) + \sum_{j=1}^{n} (w_{2ij}\mu_j(t-1)) + \cdots + \sum_{j=1}^{n} (w_{qij}\mu_j(t-q+1))\right) \tag{3}$$

Where $w_{1ij} \in [-1,+1]$, $w_{2ij} \in [-1,+1]$, ..., $w_{qij} \in [-1,+1]$ are the weights from jth node to ith node at $t, t-1, \ldots, t-q+1$ moment respectively. $\mu_j(t)$, $\mu_j(t-1)$, ..., $\mu_j(t-q+1)$ are the activation level for jth node at $t, t-1, \ldots, t-q+1$ moment respectively. Once FCM is implemented, it starts with an initial state to perform successive iteration according to Eq.3.

III. PROPOSED METHOD

The proposed novel method of fuzzy time series predicting is detailed in this section. The highlight of this method lies not only in the selection of second-order FCM predicting tools and error free triangular membership functions, but also in the fast, simple and effective way of obtaining FCM weights. Suppose the original time series $X = [x(1), x(2), \ldots, x(m)]^T$ which includes data of m time points, the block diagram of the proposed predicting method is illustrated in FIGURE.2, colorblackwhich includes three function modules: FCM representation of time series module, Solving the weights of high-order FCM module and Prediction and Defuzzification module. In what follows, function of each module in the proposed method is detailed respectively.

A. FCM REPRESENTATION OF TIME SERIES

In this module, it is mainly divided into two parts of functions, namely Normalization and Triangular function fuzzification. Normalization is to normalize the original time series, which is a typical way to standardize the data. It can unify the statistical distribution of samples and limit the preprocessed data to a certain range, such as 0 to 1. It can not only effectively eliminate the adverse effects caused by singular data, but also solve the comparability between data indicated. Consequently, normalization makes the data more suitable for comprehensive comparative evaluation. All observations $X = [x(1), x(2), \ldots, x(m)]$ are converted to $Z = [z(1), z(2), \ldots, z(m)]$ to obtain normalized time series according to Eq.(4), where $z(i)$ is the normalized value, Min is the smallest value and Max is the biggest value of the original time series X. $x(i)$ is the individual data in $X = x(1), \ldots, x(i-1), x(i), x(i+1), \ldots, x(m)$, and m is the number of data. Normalization is an effective way to simplify calculation, that is, the expression with dimension is transformed into dimensionless expression and becomes pure quantity. In particular, it can be used for comparison and weighting of indicators of different units or scales.

$$z(i) = (x(i) - Min)/(Max - Min) \quad i = 1, 2, \ldots, m. \tag{4}$$

Fuzzification refers to the process of transforming a series of input values to universe of discourse through a certain proportion and describing the input values with semantic variables. Through membership function, the membership relative to each semantic value is calculated, which is regarded as fuzzy time series. Here, the most common but simple and effective triangular membership function is considered to generate fuzzy time series data in this paper. The triangle membership functions are defined as follows:

$$\mu_j(z) = \begin{cases} (z-\delta_{j-1})/(\delta_j - \delta_{j-1}), & z \in [\delta_{j-1}, \delta_j], \\ (\delta_{j+1}-z)/(\delta_{j+1}-\delta_j), & z \in [\delta_j, \delta_{j+1}],\ j = 1, 2, \ldots, n-1. \\ 0, & otherconditions. \end{cases} \tag{5}$$

Where δ is the equidistant interval in the universe of discourse assuming [0,1] which is normalized from original data, δ_i is the ith interval. n is the number of intervals and z is the numerical value before fuzzification.

Fuzzification makes it possible to classify the normalized real data into membership vector. Assuming 6 fuzzy sets are defined, each interval of triangular membership function is a fuzzy set assigned a semantic, such as lower, low, medium, high, higher, highest etc. FIGURE.3 shows typical triangular membership functions with 6 intervals for generating membership vectors, which forms every step of the fuzzy time series. Each data after normalized belongs to every functions with different degrees. As an illustration, value z_0 has an membership vector $(0, 0, 0.4, 0.6, 0, 0)$, which means it is lower, low, medium, high, higher and highest with the possibility of 0, 0, 0.4, 0.6, 0, 0 respectively. Meanwhile, the sum of the possibilities equals to 1 definitely. The closer the degree of membership $\mu_j(z_0)$ is to 1, the higher the degree of z_0 belonging to high, and conversely, the closer $\mu_j(z_0)$ is to 0, the lower the degree of z_0 belonging to lower, low, higher or highest. It is more reasonable than classical data set theory to describe the fuzziness problem by using the membership function $\mu_j(z) \in [0, 1]$ to represent the degree of z belonging to lower, low, medium, high, higher or highest. For simplicity, when dividing the domain of a triangular membership function, the interval is set to be equal. Then the interval, namely δ, between δ_j and δ_{j-1}, $(j = 1, 2, \ldots, n-1)$, is $1/(n-1)$. Further, each $z(i)(i = 1, 2, \ldots, m)$ can be transformed into their membership expressed as $\mu(i)(i = 1, 2, \ldots, m)$ and each $\mu(i) = [\mu_1(i), \mu_2(i), \ldots, \mu_n(i)]$ is calculated according to Eq.(5) to express the degree that the normalized data belong to the semantics. In other words, each $\mu_j(i)$ can also be regarded as the activation level of the jth concept of FCM for $z(i)$. So far the fuzzy time series are generated, which consists m fuzzy subsequences $\mu(i)(i = 1, 2, \ldots, m)$ represented in form of following:

$$A = \begin{bmatrix} \mu(1) \\ \mu(2) \\ \vdots \\ \mu(m) \end{bmatrix} = \begin{bmatrix} \mu_1(1) & \mu_2(1) & \cdots & \mu_n(1) \\ \mu_1(2) & \mu_2(2) & \cdots & \mu_n(2) \\ \vdots & \vdots & \vdots & \vdots \\ \mu_1(m) & \mu_2(m) & \cdots & \mu_n(m) \end{bmatrix} = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_n \end{bmatrix}^T \tag{6}$$

where each column of A, say $V_1, V_2, \ldots, V_n$, indicates that the degree to which time series Z belongs to each semantics as time passing by. Assuming n is 6, FCM structure shown in Figure.4 is formed. $V_i(i = 1, 2, \ldots, 6)$ represent the semantic value of each concept as time passing by.

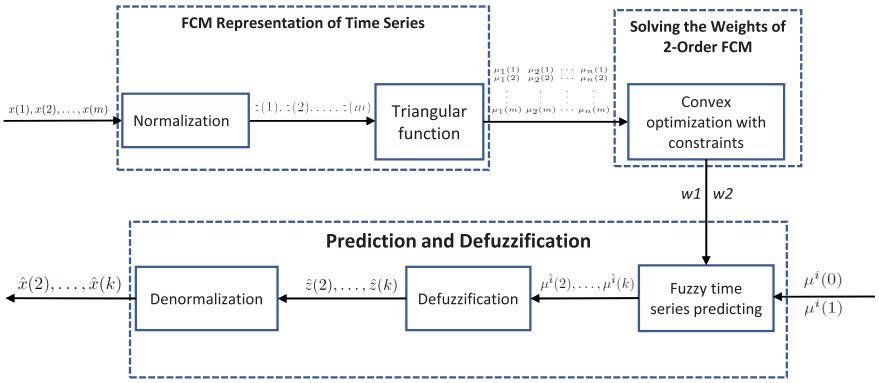


FIGURE 2. Diagram of the proposed predicting method.

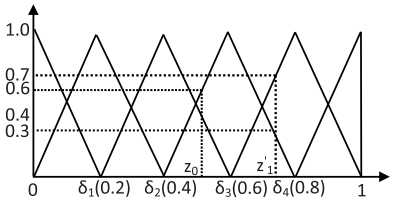


FIGURE 3. Typical triangular membership function.

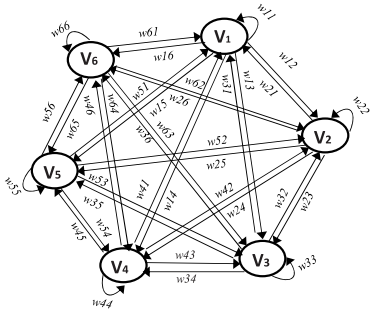


FIGURE 4. A FCM structure consisting of 6 concepts.

B. SOLVING THE WEIGHTS OF HIGH-ORDER FCM

In this module, convex optimization with constraints is introduced to find an appropriate weight matrixs which can accurately complete the iterative process when the initial value is given. For each additional order, the number of weight values to be calculated will be doubled. By comprehensive consideration, the 2-order FCM is employed to mimic the system characteristics here, which can facilitate a clearer display of the weight solving process. Further, it is assumed that the concept values at t and $t-1$ moment have almost the same effect on time $t+1$.

Because the FCM used in this algorithm is 2-ordered, the weight includes two parts: one is the weight matrix w_1 for the current time t, the other is the weight matrix w_2 for the previous time $t-1$. Here, weight matrix w_1 is $[w_{1ij}]_{\substack{i=1,2,\ldots,n\\ j=1,2,\ldots,n}}$, and w_2 is $[w_{2ij}]_{\substack{i=1,2,\ldots,n\\ j=1,2,\ldots,n}}$. Certainly both w_{1ij} and w_{2ij} are in $[-1, 1]$ for all i and j, and n is the concept number.

For the fuzzy time series A, there are n variables, c different initial state vectors $\mu^s(0) = [\mu_1^s(0), \mu_2^s(0), \ldots, \mu_n^s(0)]$ for time 0 and $\mu^s(1) = [\mu_1^s(1), \mu_2^s(1), \ldots, \mu_n^s(1)]$ for time 1, where s is $1, 2, \ldots, c$, can be considered to stimulate the FCM used in our paper within limited step k. Accordingly FCM produces c response sequences with regard to each initial state vector. Moreover, the generation of corresponding sequence is carried out step by step and lasts for k steps. In the following matrix Eq.7, $\mu^1(0), \mu^2(0), \ldots, \mu^c(0)$ are c initial vectors for time 0, and $\mu^1(1), \mu^2(1), \ldots, \mu^c(1)$ are c initial vectors for time 1, $\mu^{\hat{i}}(2), \ldots, \mu^{\hat{i}}(k), i = 2, \ldots, c$, are corresponding response vectors for initial vector $\mu^i(0), i = 1, \ldots, c$ and $\mu^i(1), i = 1, \ldots, c$, that is $\hat{d}_i$ in D displayed region by region.

Here, as the popular activation function of neural network, the monotonically increasing sigmoid shown in Eq.2 fits in with FCM. Sigmoid maps the input value between 0 and 1. The number of FCM nodes is the same as the number of semantic variables, that is, the interval number of the triangular membership function. Furthermore, the response vector of ith node ($i = 1, 2, \ldots, n$) at time $t+1$ can be produced by Eq.8.

$$D = \begin{bmatrix} \mu^1(0) \\ \mu^1(1) \\ \mu^{\hat{1}}(2) \\ \vdots \\ \mu^{\hat{1}}(k) \\ \mu^2(0) \\ \mu^2(1) \\ \mu^{\hat{2}}(2) \\ \vdots \\ \mu^{\hat{2}}(k) \\ \vdots \\ \mu^c(0) \\ \mu^c(1) \\ \mu^{\hat{c}}(2) \\ \vdots \\ \mu^{\hat{c}}(k) \end{bmatrix} = \begin{bmatrix} \mu_1^1(0) & \mu_2^1(0) & \cdots & \mu_n^1(0) \\ \mu_1^1(1) & \mu_2^1(1) & \cdots & \mu_n^1(1) \\ \mu_1^{\hat{1}}(2) & \mu_2^{\hat{1}}(2) & \cdots & \mu_n^{\hat{1}}(2) \\ \vdots & \vdots & \vdots & \vdots \\ \mu_1^{\hat{1}}(k) & \mu_2^{\hat{1}}(k) & \cdots & \mu_n^{\hat{1}}(k) \\ \mu_1^2(0) & \mu_2^2(0) & \cdots & \mu_n^2(0) \\ \mu_1^2(1) & \mu_2^2(1) & \cdots & \mu_n^2(1) \\ \mu_1^{\hat{2}}(2) & \mu_2^{\hat{2}}(2) & \cdots & \mu_n^{\hat{2}}(2) \\ \vdots & \vdots & \vdots & \vdots \\ \mu_1^{\hat{2}}(k) & \mu_2^{\hat{2}}(k) & \cdots & \mu_n^{\hat{2}}(k) \\ \vdots & \vdots & \vdots & \vdots \\ \mu_1^c(0) & \mu_2^c(0) & \cdots & \mu_n^c(0) \\ \mu_1^c(1) & \mu_2^c(1) & \cdots & \mu_n^c(1) \\ \mu_1^{\hat{c}}(2) & \mu_2^{\hat{c}}(2) & \cdots & \mu_n^{\hat{c}}(2) \\ \vdots & \vdots & \vdots & \vdots \\ \mu_1^{\hat{c}}(k) & \mu_2^{\hat{c}}(k) & \cdots & \mu_n^{\hat{c}}(k) \end{bmatrix} = \begin{bmatrix} \hat{d}_1 \\ \hat{d}_2 \\ \vdots \\ \hat{d}_m \end{bmatrix} \tag{7}$$

$$\mu_i^{\hat{s}}(t+1) = \frac{1}{1 + e^{-\lambda(\sum_{j=1}^{n} w_{1ij}\mu_j^s(t) + \sum_{j=1}^{n} w_{2ij}\mu_j^s(t-1))}} \tag{8}$$

For all i,j and s, there is a consistent one-to-one match between $\mu_i^{\hat{s}}(t+1)$ and $\mu_i^s(t+1)$ which is the actual value of the fuzzy time series at time $t+1$ fed by initial vector $\mu_i^s(0)$. In order to measure the accuracy of FCM, the degree of difference between $\mu_i^{\hat{s}}(t+1)$ and $\mu_i^s(t+1)$ is usually considered. In order to further clarify the objective of system optimization, the commonly used minimum error function is introduced into the system as shown in Eq.9.

$$\arg\min_{w} J_1 = \frac{1}{cn(k-1)\sum_{s=1}^{c}\sum_{t=1}^{k}\sum_{i=1}^{n}(\mu_i^s(t)-\mu_i^{\hat{s}}(t))^2} \tag{9}$$

where c is the overall number of initial state vectors, n is the number of nodes in FCM, k is the length of the response sequences, and t represents the tth step. Assuming the prediction error is 0 in the best case, that is, $\mu_i^{\hat{s}}(t)=\mu_i^s(t)$ for all $i(i=1,2,\ldots,n)$, then the value of J_1 will reach the minimum value of 0. To facilitate follow-up, the appearance of formula Eq.8 is transformed to Eq.10.

$$\lambda\Big(\sum_{j=1}^{c} w_{1ij}\mu_j^s(t-1)+\sum_{j=1}^{c} w_{2ij}\mu_j^s(t-2)\Big)=-\ln\Big(\frac{1}{\mu_i^{\hat{s}}(t)}-1\Big) \tag{10}$$

Since FCM starts to iterate k steps from an initial vector, the following k equations can be obtained as shown in Eq.11, according to Eq.10.

$$\begin{aligned}
\lambda\Big(\sum_{j=1}^{c} w_{1ij}\mu_j^s(1)+\sum_{j=1}^{c} w_{2ij}\mu_j^s(0)\Big)&=-\ln\Big(\frac{1}{\mu_i^{\hat{s}}(2)}-1\Big)\\
\lambda\Big(\sum_{j=1}^{c} w_{1ij}\mu_j^s(2)+\sum_{j=1}^{c} w_{2ij}\mu_j^s(1)\Big)&=-\ln\Big(\frac{1}{\mu_i^{\hat{s}}(3)}-1\Big)\\
&\vdots\\
\lambda\Big(\sum_{j=1}^{c} w_{1ij}\mu_j^s(k-1)+\sum_{j=1}^{c} w_{2ij}\mu_j^s(k-2)\Big)&=-\ln\Big(\frac{1}{\mu_i^{\hat{s}}(k)}-1\Big)
\end{aligned} \tag{11}$$

Let $-\ln\left(\frac{1}{\mu_i^{\hat{s}}(t)}-1\right)$ in Eq.11 be each element in matrix Y_i for $t=1,2,\cdots,k$; $s=1,2,\cdots,c$, let $\mu_j^s(t-1)$ be each element in matrix H_1, $\mu_j^s(t-2)$ be each element in matrix H_2. Similarly, let w_{1ij} be each element in matrix w_{1i}, and w_{2ij} be each element in matrix w_{2i}. Thus, we rewrite the above linear equations in a matrix format as shown in Eq.12.

$$\lambda(H_1w_{1i}+H_2w_{2i})=Y_i \tag{12}$$

where both H_1 and H_2 are $c(k-1)$-by-n matrix, w_{1i} and w_{2i} are n-by-1 matrix, and Y_i is an $c(k-1)$-by-1 matrix. The vector $w_{1i}^T=[w_{1i1},w_{1i2},\cdots,w_{1in}]$ and $w_{2i}^T=[w_{2i1},w_{2i2},\cdots,w_{2in}]$ are the ith row of the weight matrix w_1 and w_2 of the FCM to be learned individually. As a consequence, w_{1i} and w_{2i} can be determined by solving the constrained system of linear equations in (12).

Moreover, in consideration of the existence of errors, the above constrained linear equations usually do not have exact solutions. Therefore, going for an approximate solution that satisfies the constraints and minimizes errors $\|\ \lambda(H_1w_{1i}+H_2w_{2i})-Y_i\ \|_2$ is a viable way to solve this problem. As a result, solving Eq.12 is converted to constrained least squares problem in Eq.13.

$$\begin{aligned}
&\min: \|\lambda(\mathbf{H}_1w_{1i}+\mathbf{H}_2w_{2i})-\mathbf{Y}_i\|_2\\
&\text{s.t.}\quad \|w_{1i}\|_\infty\le 1,\quad \|w_{2i}\|_\infty\le 1,
\end{aligned} \tag{13}$$

where $\|\cdot\|_2$ is the 2-norm, $\|\cdot\|_\infty$ is the infinite norm, and $\lambda>0$ is the steepness parameter of the sigmoid function. The constraint $\|w_{1i}\|_\infty\le 1$ and $\|w_{2i}\|_\infty\le 1$ ensure that the solved weight vector w_{1i} and w_{2i} fall in the interval $[-1,1]$. Thus, the problem of solving linear equations with constraints is transformed into an optimization problem. If there is a combination of w_{1i} and w_{2i} that satisfies the inequality constraints $\|w_{1i}\|_\infty\le 1$ and $\|w_{2i}\|_\infty\le 1$, then w_{1i} and w_{2i} are feasible solutions for the optimization problem. There may be multiple feasible solutions, so the feasible solution that minimizes $\|\lambda(\mathbf{H}_1w_{1i}+\mathbf{H}_2w_{2i})-\mathbf{Y}_i\|_2$ is the optimal solution of the original linear equation system in Eq.12, where w_{1i} and w_{2i} can minimize the modeling error. Furthermore, considering improving the generalization ability of the model, i.e. regularization, and considering that sparse matrices of w_{1i} and w_{2i} of large-scale FCMs also have significant advantages in their computational efficiency, the L1 norm corresponding to the penalty term $\|w_{1i}\|_1$ and $\|w_{2i}\|_1$ shown in Eq.14 are introduced to Eq.13. Incidentally, the sparsity of w_{1i} and w_{2i} increases as β increases.

$$\begin{aligned}
&\min: \|\ \lambda(H_1w_{1i}+H_2w_{2i})-Y_i\ \|_2+\beta\ \|\ w_{1i}\ \|_1+\beta\ \|\ w_{2i}\ \|_1\\
&\text{s.t.}\quad \|\ w_{1i}\ \|_\infty\le 1, \|\ w_{2i}\ \|_\infty\le 1.
\end{aligned} \tag{14}$$

In (14), a classical convex optimization problem emerges apparently. So interior-point methods [33] such as the barrier function method, the primal-dual method, or their many enhanced versions [34], [35] can be invoked reasonably. The basic idea of the interior-point method is to continuously approach the optimal solution of weights along the search direction from an initial point. In each iteration, it solves Eq.14 to determine the search direction for the next iteration and updates the current iteration point.Benefiting from a fundamental property of convex optimization problems, any local optimal solution is also a global optimal solution. Therefore, through the interior-point method, it can help FCM find the globally optimal weights and make the system more robust, suitable for handling large-scale problems. When λ and β are hyper-parameters. Accordingly, the globally optimal solution of w_{1i} and w_{2i} are obtained.

So far, the FCM dynamic model has been established, and then, in part fuzzy time series predicting, the prediction of fuzzy time series can be completed.

Algorithm 1 Learning the Weight Matrix of Second-Ordered FCM by Way of Least Square Method.

Require:

TFtype: the type of activation function for FCM to be learned;

λ: the value of parameter of activation function corresponding to *TFtypes*, its value is larger than 0;

β: a regularization parameter, its value is larger than 0;

$x_1, x_2, \cdots, x_m$: m actual time series data used to learn FCM;

Ensure:

w_{1opt}, w_{2opt}: both w_{1opt} and w_{2opt} are well-learned n-by-n weight matrix for FCM with n concepts,w_{1opt} for current time t, w_{2opt} for previous time $t-1$

$H_2 \leftarrow \left[d_1[0:k-2,:]; d_2[0:k-2,:]; \cdots; d_m[0:k-2,:]\right]$; ▷ H_2 includes the activation vectors of FCM concepts for previous time $t-1$;

$H_1 \leftarrow \left[d_1[1:k-1,:]; d_2[1:k-1,:]; \cdots; d_m[1:k-1,:]\right]$; ▷ H_1 includes the activation vectors of FCM concepts for current time t;

$Y \leftarrow \left[d_1[2:k,:]; d_2[2:k,:]; \cdots; d_m[2:k,:]\right]$; ▷ Y includes the activation vectors of FCM concepts for the next time $t+1$; All $H_2, H_1 and Y$ are $c(k-1)-by-n$ matrices, c is the number of different initial vectors and k is simulate steps.

$i \leftarrow 0$;

while $i \leq n$ **do**

$\mathrm{Y}_i \leftarrow -\ln\left(\frac{1}{Y[:,i]} - 1\right)$;

create variables w_{1i} and w_{2i}; ▷ w_{1i} and w_{2i} both n-by-1 matrix

$variable \leftarrow w_{1i}, variable \leftarrow w_{2i}$;

$objection \leftarrow \min\{\| \lambda * (H_1 * w_{1i} + H_2 * w_{2i}) - \mathrm{Y}_i \|_2 + \| \beta * w_i \|_1\}$;

$constraint \leftarrow \{\| w_{1i} \|_\infty \leq 1, \| w_{2i} \|_\infty \leq 1\}$;

$problem \leftarrow (variable, objection, constraint)$;

$w_{1i}, w_{2i}, \leftarrow$ Solver($problem$); ▷ Solver$(\cdot)$ indicates an interior-point method based the convex optimization solver.

$w_{1opt}[i,:] \leftarrow w_{1i}^T, w_{2opt}[i,:] \leftarrow w_{2i}^T$;

$i \leftarrow i+1$;

end while

end while

C. PREDICTION AND DEFUZZIFICATION

Utilizing the methods in A and B, FCM model is derived. After that, we utilize the constructed FCM model to perform prediction according to Eq.15 in the context of given initial values. in Eq.15, $\mu_j(t)$ and $\mu_j(t-1)$ are the jth membership corresponding to the jth concept for time t and $t-1$ respectively. The predicted membership vectors are represented by$\hat{\mu}(i) = [\hat{\mu}_1(i), \hat{\mu}_2(i), \ldots, \hat{\mu}_n(i)],(i = 2, \ldots, k)$, where k is the prediction horizon. Contrary to the process of fuzzification, the predicted fuzzy membership $\hat{\mu}(i) = [\hat{\mu}_1(i), \hat{\mu}_2(i), \ldots, \hat{\mu}_n(i)], (i = 2, \ldots, k)$ is transformed to traditional time series data $\hat{Z} = [\hat{z}(2), \ldots, \hat{z}(k)]$ before denormalization according to Eq.16. Further, to restore the original time series data, it is necessary to denormalize $[\hat{z}(2), \ldots, \hat{z}(k)]$ by Eq.17. Thus, the results of predicted time series $\hat{X} = [\hat{x}(2), \ldots, \hat{x}(k)]$ are obtained to prepare for subsequent accuracy analysis. Then, the ultimate function of accuracy calculation referring to RMSE can be achieved smoothly. Here, it is worth mentioning that the triangular membership function used in this paper has the feature of the overlap level equal to $1/2$, thus fuzzification and defuzzification is error-free [43].

$$\hat{\mu}_i(t+1) = \frac{1}{1 + e^{-\lambda(\sum_{j=1}^{n} w_{1ij}\mu_j(t) + \sum_{j=1}^{n} w_{2ij}\mu_j(t-1))}} \tag{15}$$

$$\hat{z}(i) = \delta(j + \mu_{j+1}(i)) \quad \mu_j(i) > 0, \mu_{j+1}(i) > 0, \quad i = 1, 2, \ldots, m \quad j = 1, 2, \ldots, n-1 \tag{16}$$

$$\hat{x}(i) = \hat{z}(i)(Max - Min) + Min \quad i = 1, 2, \ldots, m. \tag{17}$$

The proposed method for learning weights of second-ordered FCM is summarized in Algorithm 1.

IV. EXPERIMENTAL STUDY

In this section, we implement 10 time series data sets shown in Table 1 to evaluate the effectiveness of the proposed method in this paper in two ways. One is to reveal its performance, the other is to compare with other classical approaches. In all experiments, as a feasible way to solve convex optimization problems, we draw support from "ECOS" which is presented in [34] and integrated in python package. Also the number of iteration steps is set to 100 maximally, so convergence is guaranteed. To achieve comparison with other methods, LSTM, TLSP-DE, QFCM, NFCM, Informer and IT2FCM [48], [49], [50], [54] algorithm are invoked, which are representative methods recently used in FCM weight learning. In addition, TLSP-DE adopts the same structure as the 2-order FCM in our proposed method, yet the difference is that the weight learning is implemented by differential evolution.All experiments are carried on the same computer of HP ProBook 440 G7 with CPU of Intel(R) Core(TM) i7-10510U CPU @ 1.80GHz 2.30GHz and RAM of 8.00GB. All methods use the same data set for experiments. For each data set, 80 percent of the data is applied for training, and the remaining is applied to predict.

Accordingly, the entire experiment process includes four phases for all actual time series detailed as follows.

Step 1. Prepare the training subset of FCM. Transform original time series $x_i(i = 1, 2, \ldots, m)$ into fuzzy time series $\mu_{ij}(i = 1, 2, \ldots, m; j = 1, 2, \ldots, n)$ through normalization and fuzzification according to Eq.5 and Eq.4 respectively.

Step 2. Training FCM. Carry out convex optimization method for learning weights w_{1ij} and w_{2ij} of FCM

based on the training subset. So far the FCM model is established.

Step 3. Implementing prediction on the well-learned FCM according to Eq.10. Carrying out activation level calculating by iteration, and then defuzzify and denormalizing the result sequentially according to Eq.6-7. to reconstruct numeric value.

Step 4. Validate performance. Calculating accuracy between the data generated by FCM and the actual observations to validate the performance of the well-learned FCM. Therefore, the effectiveness of the algorithm can be evaluated further.

A. 15 PUBLICLY AVAILABLE DATA SETS

15 data sets applied in this experiment include the soybean price time series, the yahoo stock time series, the sunspot time series, the Jena Climate(Tpot) time series, the Jena Climate(rho) time series, the Power consumption of Tetouan city(humidity) time series, Appliances energy prediction(Press) time series, Appliances energy prediction(power) time series, Image Recognition Task Execution Times in Mobile Edge Computing(MacPro1) time series and Condition monitoring of hydraulic systems(Cooling efficiency) time series, Appliances energy prediction(Press) long time series, Appliances energy prediction(power) long time series, the Power consumption of Tetouan city(humidity) long time series, the Jena Climate(rho) long time series and the Jena Climate(Tpot) long time series. The specific information for each data set is shown in Table 1. Considering the regularity and irregularity of data changes in time series, various types of time series are selected in the experiment, including trending time series, seasonal time series, stochastic time series, and comprehensive time series. Among them, trending time series present a relatively slow and long-term continuous upward, downward, or stable movement trend, but the magnitude of the change may not be equal like the Jena Climate(Tpot) time series, the Jena Climate(rho) time series and Appliances energy prediction(Press) time series. The regular pattern of seasonal time series with peaks and troughs that alternate with natural seasons or certain cycles like the sunspot time series, Appliances energy prediction(power) time series, the Jena Climate(rho) long time series and the Jena Climate(Tpot) long time series. The stochastic time series vary individually and exhibit statistical regularity as a whole like the yahoo stock time series, the Power consumption of Tetouan city(humidity) time series, the Image Recognition Task Execution Times in Mobile Edge Computing(MacPro1) time series, the Appliances energy prediction(Press) long time series, the Appliances energy prediction(power) long time series and the Power consumption of Tetouan city(humidity) long time series. A comprehensive time series is the superposition or combination of various changes like the soybean price time series and Condition monitoring of hydraulic systems(Cooling efficiency) time series.The trend for each dataset is shown in Figure.5. Data set soybean price records the prices of soybean in guangdong from January 4, 2010 to June 6, 2014. Data set yahoo stock records the finance yahoo stock data(high point) from September 28, 2010 to September 25, 2015. Data set sunspots records 13-month smoothed monthly total sunspot number. Data set jena climate from Max Planck Institute for Biogeochemistry includes 14 characteristic indicators such as temperature and gas density we used in experiments. Tetouan city from UCI is power consumption of Tetouan city data set which is related to power consumption, humidity of Tetouan city. Here, humidity data is selected as training and predicting time series. Appliances energy from UCI is appliances energy prediction data set which includes several characteristic indicators such as pressure and power consumption in a low energy building. Image recognition times from UCI is image recognition task execution times in mobile edge computing data set, which records task execution times for four edge servers submitted by edge node. Here, task execution times of MacBookPro1 is selected as training and predicting time series. Hydraulic systems cooling efficiency from UCI is condition monitoring of hydraulic systems data set, in which four fault types are superimposed with several severity grades impeding selective quantification. Here, cooling efficiency is selected as training and predicting time series.

B. PERFORMANCE METRICS

To evaluate the performance of well-learned FCMs, several performance metrics such as In-sample error(J_1),Out-of-sample error(J_2),Model error(J_3),Mean of sensitivity and specificity(J_4) etc. are usually used. In fact, J_3 and J_4 cannot be calculated, because in real world problems the weight matrix is unknown. Compared with J_1, J_2 is more suitable for evaluating the generalization ability of FCM model. it measures the difference between the predicted data and the actual ones. The smaller the value of J_2, the stronger the generalization ability of the well-learned FCM model by cross validation is, which is exactly what we expect. Common performance metrics J_2 mainly include: MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error), MSE (Mean Square Error) and RMSE (Root Mean Square Error). All MAE, MAPE, MSE and RMSE can be used to measure the error between predicted and true data [47], [48], [49], [50]. MAE is the absolute difference with a linear relationship between loss and error, and is the simplest regression measure. But MAE uses a modulus function which is not differentiable at all points, so it has certain limitations as a loss function. MAPE refers to the mean absolute percentage error, which is a relative measure sensitive to errors and suitable for problems with large dimensional differences. But when an actual data is 0, MAPE cannot be calculated. MSE is mean square error, so the relationship between MSE and error is square. Due to the square operation on the difference value, larger error values have a greater impact on the fit, which helps to more sensitively capture the prediction error of the model. Moreover, the square function is differentiable at all points, so it can be used as a loss function. But the unit of MSE is the square of the actual data. RMSE is the square

TABLE 1. Brief introduction to 15 data sets.

Number	Data Set	Description	Train set	Test set	Source	Publicly avaliable
1	Soybean Price	the prices of soybean in guangdong from January 4, 2010 to June 6, 2014	1080	271	$https://download.csdn.net/download/qq_40797107/10477157$	Yes
2	Yahoo stock	recording the finance yahoo stock data(high point) from September 28, 2010 to September 25,2015	1007	251	$https://download.csdn.net/download/weixin_42126677/18402174$	Yes
3	sunspots	records 13-month smoothed monthly total sunspot number	2632	658	$https://www.sidc.be/silso/datafiles$	Yes
4	jena climate(Tpot)	temperature from Max Planck Institute for Biogeochemistry	852	213	$https://blog.csdn.net/m063642362/article/details/127721083$	Yes
5	jena climate(rho)	gas density from Max Planck Institute for Biogeochemistry	2188	547	$https://blog.csdn.net/m063642362/article/details/127721083$	Yes
6	Tetouan city(humidity)	humidity of Tetouan city	2000	499	http://archive.ics.uci.edu/	Yes
7	Appliances energy(Press)	pressure in a low energy building	1759	440	http://archive.ics.uci.edu/	Yes
8	Appliances energy(Power)	power consumption in a low energy building	1280	320	http://archive.ics.uci.edu/	Yes
9	Image Recognition Times(Mac1)	image recognition task execution times in mobile edge computing of MacBookPro1	800	200	http://archive.ics.uci.edu/	Yes
10	Hydraulic systems Cooling efficiency	condition monitoring of hydraulic systems for cooling efficiency	1764	441	http://archive.ics.uci.edu/	Yes
11	Appliances energy(Press) long	pressure in a low energy building	15788	3947	http://archive.ics.uci.edu/	Yes
12	Appliances energy(Power) long	power consumption in a low energy building	15788	3947	http://archive.ics.uci.edu/	Yes
13	Tetouan city(humidity) long	humidity of Tetouan city	41932	10483	http://archive.ics.uci.edu/	Yes
14	jena climate(rho) long	gas density from Max Planck Institute for Biogeochemistry	56916	14229	$https://blog.csdn.net/m063642362/article/details/127721083$	Yes
15	jena climate(Tpot) long	temperature from Max Planck Institute for Biogeochemistry	80000	20000	$https://blog.csdn.net/m063642362/article/details/127721083$	Yes

root of MSE, which not only has the advantages of MSE, but also solves the unit problem of MSE, i.e. its unit will be the same as the actual data. In view of this, RMSE shown in Eq.18 is chosen to evaluate the accuracy of FCMs in predicting out-of-sample values as used in recent literature on evaluating the performance of fuzzy cognitive maps [51], [52], [53]. FCM with lower RMSE for the out-of-sample data will have lower prediction error. Further, it will have the expectation of better predictive power in the future. Therefore, in this paper we chooses RMSE as the most appropriate performance metric of J_2.

$$RMSE = \sqrt{\frac{1}{k}\sum_{t=1}^{k}(x_t - \hat{x}_t)} \tag{18}$$

where k is the predicting steps, namely prediction horizon, x_t and $\hat{x}_t$ stand for the actual observation value and predicted ones at time t, respectively.

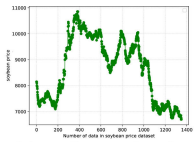
C. DEMONSTRATION OF THE PROPOSED METHOD

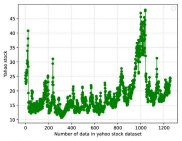
The soybean price date set is first used to validate the performance of the well-learned FCM as the demonstration. In accordance with the performance, the feasibility and effectiveness of the proposed method in this paper present evidently.

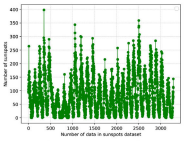
Specially taking $n = 20$, $\lambda = 3.0$, $\beta = 0.0004$ as an example, the process of implementing the method proposed is described in detail as follows.

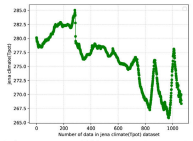
1) FUZZIFICATION

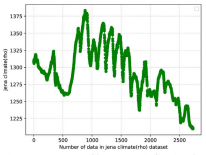
In this demonstration example, 20 fuzzy sets are divided. Based on 20 fuzzy sets, the original time series is fuzzified. The first 80% of the soybean price time series are used as training set, and the remaining 20% are used as test set to validate the prediction accuracy of FCM constructed by the proposed method.

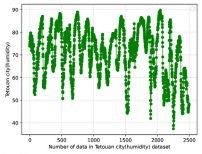

(a) soybean price time series

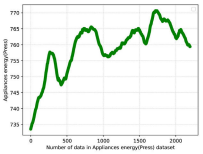

(b) yahoo stock time series

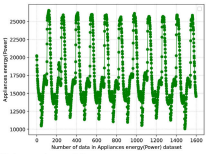

(c) sunspot time series

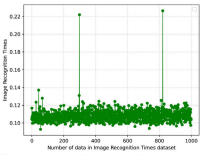

(d) Jena Climate(Tpot) time series

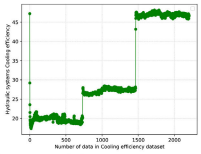

(e) Jena Climate(rho) time series

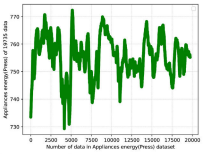

(f) Power consumption of Tetouan city(humidity) time series

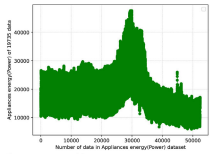

(g) Appliances energy prediction(Press) time series

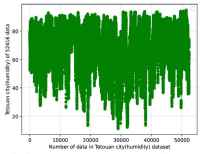

(h) Appliances energy prediction(power) time series

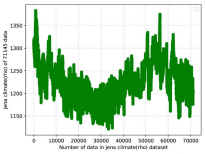

(i) Image Recognition Task Execution Times in Mobile Edge Computing(MacPro1) time series


(j) Condition monitoring of hydraulic systems(Cooling efficiency) time series


(k) Appliances energy prediction(Press) long time series


(l) Appliances energy prediction(power) long time series


(m) Power consumption of Tetouan city(humidity) long time series


(n) Jena Climate(rho) long time series

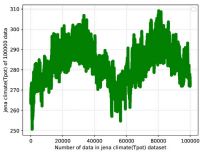

(o) Jena Climate(Tpot) long time series

FIGURE 5. Trends of values in 15 publicly available data sets.

First of all, the concepts of soybean price data set are extracted in the form of numerical value. Where the triangular function is applied to transform the original numeric time series into the fuzzy time series with semantic prototypes without loss of accuracy. Each fuzzy set also known as concept can be exhibited in the form of a certain fuzzy semantics. Here, because the number of FCM concepts is 20, the corresponding fuzzy semantics is to divide the price into 20 levels from low to high as below. Simultaneously the fuzzy time series are expressed by membership matrix which voices the fuzzy relationships between each data sample and concepts of FCM individually.

- the price of soybean is lowest, where locate around 6908;
- the price of soybean is second lowest, where locate around 7115;
- ⋮
- the price of soybean is medium, where locate around 8775;
- ⋮
- the price of soybean is second highest, where locate around 10642;
- the price of soybean is highest, where locate around 10850;

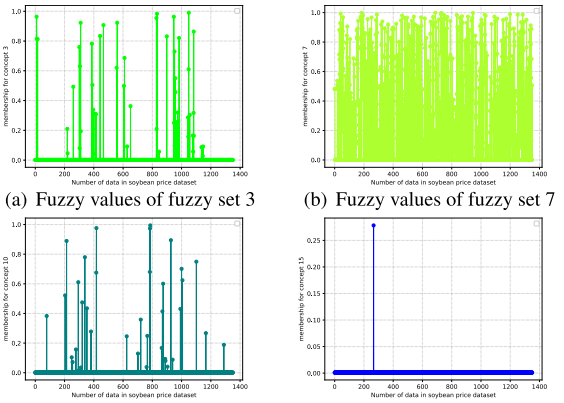
After normalization and fuzzification according to Eq.4 and Eq.5 respectively, original time series shown in Figure.5(a) are transformed to fuzzy time series consisting of 20 fuzzy sets. In order to further demonstrate the appearance of fuzzy time series, Figure.6 shows fuzzy values of 4 fuzzy sets randomly.

2) SOLVING FCM WEIGHTS

After obtaining the fuzzy time series, it is urgent to adopt appropriate methods to obtain FCM weights. Here, the interior-point method is borrowed from convex optimization to learn parameters w_1 and w_2 of second-ordered FCM according to the fuzzy time series data $\mu_i(i = 1, 2, \ldots, 20)$. The parameter w_1 and w_2 are shown in Figure.7. Incidentally, the reason for selecting a 20 node FCM here is to fully display the weight matrix.

3) PREDICTION PERFORMANCE

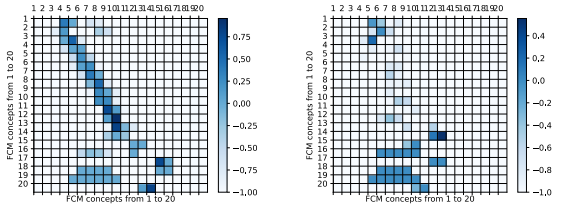
So far, 2-order FCM modeling has been implemented and then prediction can be accomplished for the latter 20% test data according to Eq.15. Subsequently, defuzzifying the predicted $\hat{\mu})(i), (i = 1, \ldots, 271)$ to $\hat{z}(i), (i = 1, \ldots, 271)$ by Eq.16, and denormalizing $\hat{z}(i)$ to numeric value $\hat{x}(i), (i = 1, \ldots, 271)$ by Eq.17. Finally, RMSE 24.131 is calculated to validate the prediction performance of the method proposed in this paper. FIGURE.8 shows the original soybean price time series and the predicted results at $n = 20$, $\lambda = 3.0$ and $\beta = 0.0004$.



(a) Fuzzy values of fuzzy set 3 (b) Fuzzy values of fuzzy set 7

(c) Fuzzy values of fuzzy set 10 (d) Fuzzy values of fuzzy set 15

FIGURE 6. Fuzzy values of 4 fuzzy sets.



(a) w_1 (b) w_2

FIGURE 7. w_1 and w_2 for 2-order FCM with 20 concepts.

D. IMPACT OF VARIOUS PARAMETERS ON PREDICTION PERFORMANCE

Based on the experiment process described above, the prediction performance is verified through many experiments, and the influence of different parameters on the FCM prediction results is shown as well. In detail, the whole dataset with 1351 samples is divided into two parts. The first 1080 samples are used as the training set, and the remaining 271 samples are performed as test set. Table 2 reports the experimental results measured by RMSE in condition of different parameters of FCM concept number n,steepness parameter λ and regularization parameter β. Where the values of n are in the range [20,200], the values of λ are in the range [2.0,5.0) with a step of 0.5, and the values of β are in the range [0.0001,0.0005) with a step of 0.0001. FIGURE 9 and FIGURE 10 exhibits the plots of the variety of performance RMSE at different values of parameters n,λ and β.

According to Table 2, the maximum RMSE is 24.131 with n is 20,λ is 3.0 and β is 0.0004. Meanwhile the minimum one is 20.232 with n is 200,λ is 3.5 and β is 0.0001. Therefore we can achieve better prediction accuracy by adjusting the values of parameter n, λ, and β. In FIGURE 9 and FIGURE 10,the influence trend of individual parameter on RMSE is shown as illustrations, which can be used as an effective reference while optimizing parameters. FIGURE 9 shows that in the case of fixing number of FCM concepts n,such as $n = 40, 100, 150, 200$, the RMSE values exhibit a downtrend when λ moves from low to high. The more the steepness parameter λ is, the smaller the value of RMSE is, and the other way around. FIGURE 11 demonstrates the impact of various parameters on RMSE further. FIGURE 11(a) shows

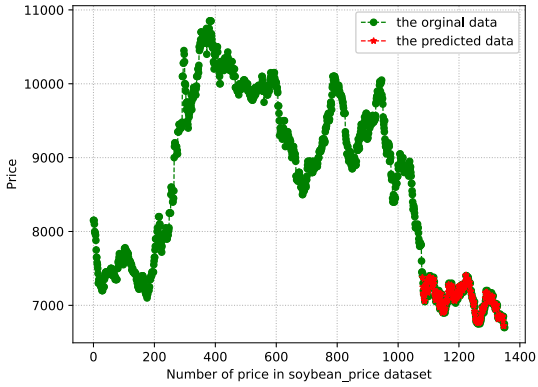


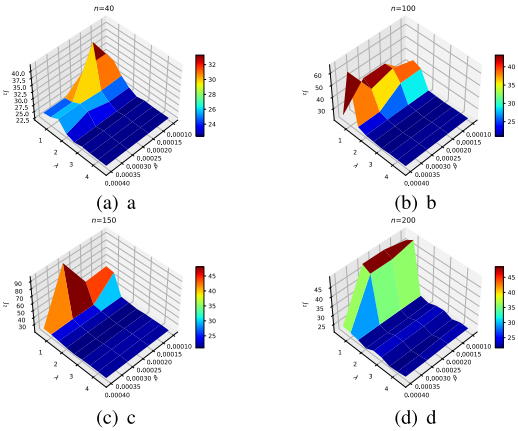
FIGURE 8. The original data and predicted data of soybean price time series($n = 20$, $\lambda = 3.0$, $\beta = 0.0004$).

the variation pattern of RMSE with a fixed n value and an increase in λ. As λ increases, RMSE shows a significant change from high to low. And when λ increases to a certain value, such as after 2.0, it will maintain a relatively stable low value. The explanation for this is that the steeper the sigmoid as an activation function is, the more sensitive the FCM is to changes in data, which means that the sensing ability of the FCM is more flexible. However, when λ reaches a certain value, the perception ability of FCM will remain stable and no longer show significant changes. Similarly, FIGURE 10 shows that in the case of fixing value of steepness parameter λ, such as $\lambda = 3.0, 3.5, 4.0, 4.5$, the RMSE values exhibit an downtrend when n moves from low to high, although RMSE may fluctuate slightly as n increases. FIGURE 11(b) shows this change pattern more clearly. Generally, the more the FCM concepts are, the smaller the value of RMSE is, and the other way around. It implies that for small-scale FCMs, it is barely enough to reflect the dynamic characteristics of large-scale fuzzy time series. As the scale of FCM increases, the dynamic characteristics of the system are effectively embodied, so RMSE decreases. Yet as FCM concepts number n and the steepness parameter λ are fixed, RMSEs are less sensitive to β as further shown in FIGURE 11(c) and FIGURE 11(d). FIGURE 11(c) shows the variation of RMSE with increasing β when λ is fixed and n is at different values. FIGURE 11(d) shows the variation of RMSE with increasing β when n is fixed and λ is at different values. Both FIGURE 11(c) and FIGURE 11(d) exhibit a common characteristic, which is that β has a very weak impact on RMSE. That's because the biggest role of β is not to affect prediction accuracy, but to adjust the number of zeros in the weight matrix.

Here, it is worth mentioning that FCM concept number n, steepness parameter λ and regularization parameter β are all given in the form of hyper-parameters so far. The consequence is that the predictive performance obtained in the experiment may not be optimal sufficiently. A grid optimization strategy can be used to create a grid of possible values for n, λ and β. Each iteration attempts a combination of these hyper-parameters in a specific order. It fits the model on

TABLE 2. The experimental result of RMSE under FCM concept number n, steepness parameter λ and regularization parameter β for soybean price of Guangdong from January 4, 2010 to June 6, 2014.

n	λ	RMSE with different β			
		0.0001	0.0002	0.0003	0.0004
20	2.0	23.478	23.482	23.483	23.484
	2.5	23.908	23.91	23.912	23.914
	3	24.124	24.127	24.129	24.131
	3.5	23.91	23.911	23.913	23.915
	4	23.901	23.902	23.905	23.907
	4.5	23.886	23.901	23.902	23.903
40	2.0	22.618	22.617	22.615	22.362
	2.5	22.235	22.327	22.327	22.325
	3	22.395	22.393	22.391	22.391
	3.5	22.414	22.414	22.413	22.413
	4	22.331	22.33	22.33	22.33
	4.5	22.229	22.229	22.229	22.229
60	2.0	22.158	22.197	22.198	22.203
	2.5	21.808	21.808	21.671	21.571
	3	21.697	21.698	21.698	21.698
	3.5	21.355	21.355	21.355	21.356
	4	21.585	21.586	21.586	21.586
	4.5	21.353	21.354	21.354	21.353
100	2.0	21.673	21.666	21.679	21.753
	2.5	21.335	21.334	21.331	21.24
	3	21.045	21.152	21.239	21.248
	3.5	21.601	21.601	21.582	21.311
	4	21.513	21.513	21.513	21.513
	4.5	21.517	21.518	21.518	21.518
150	2.0	21.117	21.116	21.125	21.102
	2.5	20.498	20.498	20.497	20.499
	3	21.247	21.281	21.284	21.284
	3.5	21.548	21.548	21.549	21.566
	4	21.586	21.734	21.733	21.733
	4.5	21.555	21.553	21.556	21.559
200	2.0	20.817	20.863	20.74	20.794
	2.5	21.43	21.429	21.429	21.427
	3	21.61	21.139	21.136	21.134
	3.5	20.232	20.259	20.26	20.26
	4	20.891	20.89	20.909	20.909
	4.5	20.864	20.862	20.882	20.881



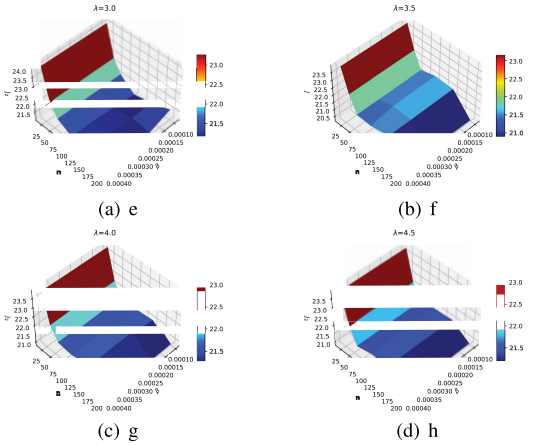
(a) a (b) b (c) c (d) d

FIGURE 9. Plots of the values of performance of RMSE under parameter λ and β with fixed concept number n for soybean price of guangdong from January 4, 2010 to June 6, 2014.

every possible combination of hyper-parameters, and records the performance of the model. Finally, it returns the best model with the most suitable n, λ and β.

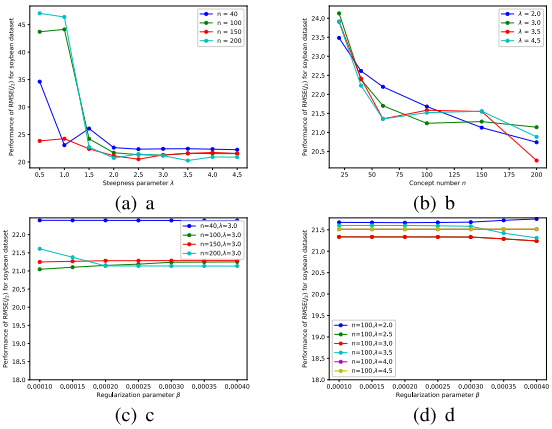
E. COMPARISON OF PERFORMANCE WITH OTHER METHODS

In order to compare the FCM modeling and prediction method(namely TLSP) proposed in this paper with other classical methods, and to evaluate the robustness and wide applicability of TLSP, multiple data sets from different fields have been introduced. Table 3 reports the comparison results of the proposed prediction method TLSP with other methods [25], [26], [27], [28]. The performances of TLSP on datasets soybean price has been shown in detail in section IV-C. Sequentially, experiments on another 9 both



(a) e (b) f (c) g (d) h

FIGURE 10. Plots of the values of performance of RMSE under parameter n and β with fixed steepness parameter λ for soybean price of guangdong from January 4, 2010 to June 6, 2014.



(a) a (b) b (c) c (d) d

FIGURE 11. The impact of parameters λ, n and β on the prediction accuracy for soybean price dataset.

public and real datasets show that TLSP works better at prediction supported by target RMSE. For all data sets, the first 80% is used as the training set, and the rest is used as the test set.

The rows in Table 3 list 15 experimental data sets, and the columes show the performance RMSE generated by method TLSP, LSTM, TLSP-DE, QFCM, NFCM, Informer and IT2FCM respectively. The experiment results in Table 3 reveal that TLSP can obtain the minimum RMSE value, demonstrate satisfactory prediction performance for all 15 data sets. In addition, Table 3 also shows the computation time required for each method to establish the corresponding model, represented by T_s. The T_s of method TLSP is superior to other methods listed. Overall, under the premise of automatically acquiring the relationships among FCM concepts by convex optimization, and accomplishing prediction without human intervention, the proposed method TLSP can not only get higher prediction accuracy, but also has superiority in computation time. The reasons for this result are mainly attributable to the advantages and disadvantages of various methods. LSTM, NFCM and Informer are essentially neural network based algorithms that can effectively capture nonlinear features in large-scale time series, leading to

TABLE 3. Experimental results obtained by respectively using the proposed algorithm TLSP, LSTM, TLSP-DE, QFCM, NFCM, Informer and IT2FCM.

Data Set	TLSP		LSTM		TLSP-DE		QFCM		NFCM		Informer		IT2FCM	
	$RMSE$	T_s	$RMSE$	T_s	$RMSE$	T_s	$RMSE$	T_s	$RMSE$	T_s	$RMSE$	T_s	$RMSE$	T_s
Soybean Price	20.49	105.7	79.72	2998.4	57.42	611.2	51.33	421.6	42.51	876.5	37.89	1058.3	49.71	531.5
Yahoo stock	0.54	141.3	1.61	3157.4	1.80	724.6	0.85	889.7	1.02	1740.6	0.97	2123.5	2.04	942.6
sunspots	9.35	123.1	11.31	3182.3	10.54	756.9	11.03	1233.4	9.52	2426.4	13.41	1755.9	10.23	678.4
jena climate(Tpot)	0.15	195.4	0.19	3315.8	0.21	788.9	0.16	593.2	0.19	1047.6	0.19	978.9	0.18	623.3
jena climate(rho)	0.38	167.8	0.39	3395.1	0.42	824.6	0.43	724.6	0.39	2762.4	0.38	1265.4	0.40	812.7
Tetouan city(humidity)	0.70	83.8	0.88	2732.4	0.84	693.4	0.86	713.8	0.92	1257.6	0.76	1568.8	0.76	688.8
Appliances energy(Press)	0.016	101.2	0.017	3266.3	0.017	778.9	0.025	846.4	0.018	2172.4	0.019	1871.3	0.017	763.9
Appliances energy(Power)	115.55	65.3	118.43	2443.7	118.75	655.7	117.26	224.5	147.31	1049.6	122.53	1535.5	126.54	358.6
Image Recognition Times(Mac1)	0.0036	243.4	0.0041	4381.2	0.0041	856.3	0.0044	387.3	0.0037	2723.5	0.0039	1946.5	0.0038	572.4
Hydraulic systems Cooling efficiency	0.16	187.3	0.21	3022.6	0.19	755.6	0.18	221.6	0.29	1887.8	0.17	1636.3	0.17	821.2
Appliances energy(Press) long	0.018	1247	0.020	34145	0.019	11246	0.023	13067	0.021	28743	0.021	20742	0.019	11347
Appliances energy(Power) long	113.24	806	117.54	27329	118.16	9426	116.34	3153	142.98	14536	119.47	17625	124.72	52714
Tetouan city(humidity) long	0.74	1165	0.94	33017	0.89	9872	0.82	9825	0.88	15978	0.75	18223	0.76	83217
jena climate(rho) long	0.36	1872	0.38	39182	0.49	12337	0.51	9887	0.39	30734	0.41	15374	0.44	11324
jena climate(Tpot) long	0.14	2253	0.21	38463	0.25	10867	0.15	8135	0.23	15488	0.22	12971	0.17	8194

high prediction accuracy. However, these methods require computing neural network weights, which is computationally intensive and time-consuming, and leads to a preference for higher T_s. For TLSP-DE, QFCM, NFCM and IT2FCM, they applied FCM as the core structure. Fortunately, FCM can effectively simulate the intrinsic connections of complex systems, effectively explain and predict system behavior, and has strong knowledge reasoning capabilities. However, TLSP-DE, QFCM and IT2FCM are essentially population based algorithms. Although they have the advantage of fast convergence speed, they cannot guarantee convergence to the global optimal solution. Although TLSP-DE, QFCM, NFCM and IT2FCM possess the same 2-order FCM structure as TLSP and also achieved fairly good predictive performance, TLSP still outperforms them in terms of prediction accuracy. The reason for the better prediction accuracy of TLSP is due to the adoption of convex optimization that converges to the optimal weight solution, and both fuzzification and defuzzification with error-free. Moreover, the computational complexity of TLSP is relatively small, so T_s is relatively short. As a whole, for typical time series, i.e. trending time series, seasonal time series, stochastic time series, and comprehensive time series, TLSP exhibits excellent performance both in terms of prediction accuracy and computation time.

In order to further demonstrate the impact of the order of FCM on prediction performance, Table 4 lists RMSE and T_s of the TLSP method for predicting 10 time series using 2-order FCM, 5-order FCM, 10-order FCM, 20-order FCM and 50-order FCM, respectively. In theory, as the order of FCM increases, the consideration of historical information becomes more sufficient, and the prediction accuracy will become higher. However, as can be seen from Table 4, as the order of FCM increases, the prediction accuracy does show a certain trend of improvement at first. As FCM reaches a certain order, the prediction accuracy does not improve, but instead decreases. Analyzing the reason, it is assumed in the experiment that various historical data has the same impact on prediction. While in reality, the older the historical data, the smaller the impact on prediction is more likely. Choosing appropriate weights for various historical data is particularly important and challenging. For T_s, as the order

TABLE 4. Experimental results obtained by respectively using the proposed algorithm TLSP with 2-order FCM, 5-order FCM, 10-order FCM, 20-order FCM and 50-order FCM.

Data Set	2-order FCM		5-order FCM		10-order FCM		20-order FCM		50-order FCM	
	$RMSE$	T_s	$RMSE$	T_s	$RMSE$	T_s	$RMSE$	T_s	$RMSE$	T_s
Soybean Price	20.49	105.7	18.23	114.6	17.22	128.5	16.17	138.5	31.24	188.7
Yahoo stock	0.54	141.3	0.51	155.7	0.49	172.3	0.53		0.66	196.2
sunspots	9.35	123.1	9.24	128.4	9.24	137.9	9.33	144.7	9.89	165.8
jena climate(Tpot)	0.15	195.4	0.14	199.7	0.15	216.4	0.15	219.3	0.17	244.6
jena climate(rho)	0.38	167.8	0.37	171.3	0.35	185.4	0.37	194.6	0.45	221.4
Tetouan city(humidity)	0.70	83.8	0.68	85.2	0.69	88.3	0.71	93.2	0.81	107.9
Appliances energy(Press)	0.016	101.2	0.015	105.3	0.015	111.7	0.017	127.8	0.022	156.1
Appliances energy(Power)	115.55	65.3	112.2	68.6	109.4	73.1	119.8	80.5	134.4	89.3
Image Recognition Times(Mac1)	0.0036	243.4	0.0034	249.9	0.0036	255.4	0.0042	260.1	0.0048	277.3
Hydraulic systems Cooling efficiency	0.16	187.3	0.15	192.3	0.14	201.2	0.17	209.3	0.21	222.3
Appliances energy(Press) long	0.018	1247	0.014	1286	0.015	1354	0.019	1567	0.024	1889
Appliances energy(Power) long	113.24	806	109.3	879	110.4	927	121.6	998	128.5	1117
Tetouan city(humidity) long	0.74	1165	0.64	1274	0.67	1369	0.70	1457	0.79	1762
jena climate(rho) long	0.36	1872	0.36	1984	0.33	2166	0.36	2543	0.48	2837
jena climate(Tpot) long	0.14	2253	0.13	2268	0.13	2374	0.12	2426	0.19	2816

of FCM increases, the computational load will also increase and the computational time will be extended.

V. CONCLUSION

In this paper, an effective data-driven prediction method for long-term or complex time series is proposed. The proposed method is based on high-order FCM, and the key to constructing FCM is the learning of weights among FCM concepts. By analyzing dynamic characteristics, this paper convert the acquisition of weights into a constrained convex optimization problem, which can promptly learn weights, especially for large-scale FCMs with high quality. The experimental process of this method includes the following four steps in order: (1) Preparing the training subset of FCM through normalization and triangular fuzzification. (2) Training FCM by solving least square problem. (3) Implementing prediction on the well-learned FCM. (4) Validating performance by RMSE. Fifteen public available benchmark time series: the soybean price time series, the yahoo stock time series, the sunspot time series, the Jena Climate(Tpot) time series, the Jena Climate(rho) time series, the Power consumption of Tetouan city(humidity) time series, Appliances energy prediction(Press) time series, Appliances energy prediction(power) time series, Image Recognition Task Execution Times in Mobile Edge Computing(MacPro1) time series, Condition monitoring of hydraulic systems(Cooling efficiency) time series, Appliances energy prediction(Press) long time series, Appliances energy prediction(power) long time series, the Power consumption of Tetouan city(humidity) long time series, the Jena Climate(rho) long time series and the Jena Climate(Tpot) long time series are used to validate effectiveness and feasibility of the proposed method. The results of experiments show that the proposed prediction method can get weights conveniently, moreover better prediction accuracy and prediction time are achieved. Compared to neural network-based algorithms and population-based algorithms, the proposed method not only significantly avoids a large amount of computation, but also obtains the global optimal weights. Therefore, the proposed method ensures high prediction accuracy while shortening computation time. Further, by analyzing the details of prediction results of the benchmark time series, several interesting conclusions can be drawn about the impacts of FCM concepts number n, steepness index λ and regularization parameter β. However, this method still has relatively weak points, mainly reflected in two aspects.

On one hand, despite multiple combinations of parameters being provided, all current parameters are given in the form of hyper-parameters without sufficient optimization. On the other hand, the impact of FCM order, i.e. various historical data on prediction performance needs to be further optimized. As such, in the future, it can be considered to utilize further parameter optimization and optimized higher-order FCM to enhance the prediction precision. Additionally, merging with prior knowledge while extracting FCM concepts is a promising direction of future investigation.

REFERENCES

[1] E. I. Papageorgiou and J. L. Salmeron, "A review of fuzzy cognitive maps research during the last decade," *IEEE Trans. Fuzzy Syst.*, vol. 21, no. 1, pp. 66–79, Feb. 2013.

[2] B. Kosko, "Fuzzy cognitive maps," *Int. J. Man-Mach. Stud.*, vol. 24, no. 1, pp. 65–75, Jan. 1986.

[3] R. Axelrod, *Structure of Decision: The Cognitive Maps of Political Elites*, vol. 72, no. 3. Princeton, NJ, USA: Princeton Univ. Press, 2015, pp. 10–17.

[4] C. D. Stylios and P. P. Groumpos, "Modeling complex systems using fuzzy cognitive maps," *IEEE Trans. Syst. Man, Cybern. A, Syst. Humans*, vol. 34, no. 1, pp. 155–162, Jan. 2004.

[5] W. Pedrycz, A. Jastrzebska, and W. Homenda, "Design of fuzzy cognitive maps for modeling time series," *IEEE Trans. Fuzzy Syst.*, vol. 24, no. 1, pp. 120–130, Feb. 2016.

[6] G. Nápoles, I. Grau, R. Bello, and R. Grau, "Two-steps learning of fuzzy cognitive maps for prediction and knowledge discovery on the HIV-1 drug resistance," *Expert Syst. Appl.*, vol. 41, no. 3, pp. 821–830, Feb. 2014.

[7] W. Froelich and J. L. Salmeron, "Evolutionary learning of fuzzy grey cognitive maps for the forecasting of multivariate, interval-valued time series," *Int. J. Approx. Reasoning*, vol. 55, no. 6, pp. 1319–1335, Sep. 2014.

[8] S. Mei, Y. Zhu, X. Qiu, X. Zhou, Z. Zu, A. V. Boukhanovsky, and P. M. A. Sloot, "Individual decision making can drive epidemics: A fuzzy cognitive map study," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 2, pp. 264–273, Apr. 2014.

[9] S. Lee, J. Yang, and J. Han, "Development of a decision making system for selection of dental implant abutments based on the fuzzy cognitive map," *Expert Syst. Appl.*, vol. 39, no. 14, pp. 11564–11575, Oct. 2012.

[10] E. I. Papageorgiou, "Learning algorithms for fuzzy cognitive maps—A review study," *IEEE Trans. Syst. Man, Cybern. C, Appl. Rev.*, vol. 42, no. 2, pp. 150–163, Mar. 2012.

[11] W. Stach, L. Kurgan, W. Pedrycz, and M. Reformat, "Genetic learning of fuzzy cognitive maps," *Fuzzy Sets Syst.*, vol. 153, no. 3, pp. 371–401, Aug. 2005.

[12] W. Stach, L. Kurgan, and W. Pedrycz, "A divide and conquer method for learning large fuzzy cognitive maps," *Fuzzy Sets Syst.*, vol. 161, no. 19, pp. 2515–2532, Oct. 2010.

[13] J. Liu, Y. Chi, and C. Zhu, "A dynamic multiagent genetic algorithm for gene regulatory network reconstruction based on fuzzy cognitive maps," *IEEE Trans. Fuzzy Syst.*, vol. 24, no. 2, pp. 419–431, Apr. 2016.

[14] X. Zou and J. Liu, "A mutual information-based two-phase memetic algorithm for large-scale fuzzy cognitive map learning," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 4, pp. 2120–2134, Aug. 2018.

[15] E. I. Papageorgiou and P. Groumpos, "A new hybrid learning algorithm for fuzzy cognitive maps learning," *Appl. Soft Comput.*, vol. 5, no. 4, pp. 409–431, 2005.

[16] Y. Zhu and W. Zhang, "An integrated framework for learning fuzzy cognitive map using RCGA and NHL algorithm," in *Proc. 4th Int. Conf. Wireless Commun., Netw. Mobile Comput.*, Oct. 2008, pp. 1–5.

[17] X. Liu and Y. Zhang, "Numerical dynamic modeling and data driven control via least square techniques and Hebbian learning algorithm," *Int. J. Numer. Anal. Model.*, vol. 7, no. 1, pp. 66–86, 2010.

[18] Q. Song and B. S. Chissom, "Fuzzy time series and its models," *Fuzzy Sets Syst.*, vol. 54, no. 3, pp. 269–277, Mar. 1993.

[19] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series—Part I," *Fuzzy Sets Syst.*, vol. 54, no. 1, pp. 1–9, Feb. 1993.

[20] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series—Part II," *Fuzzy Sets Syst.*, vol. 62, no. 1, pp. 1–8, Feb. 1994.

[21] L.A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, pp. 338–353, Sep. 1965.

[22] R. E. Ismail, "Modied weighted for enrollment forecasting based on fuzzy time series," *MATEM ATIKA*, vol. 25, no. 1, pp. 67–78, 2009.

[23] R. Efendi, Z. Ismail, and M. M. Deris, "A new linguistic out-sample approach of fuzzy time series for daily forecasting of Malaysian electricity load demand," *Appl. Soft Comput.*, vol. 28, pp. 422–430, Mar. 2015.

[24] R. Efendi, M. M. Deris, and Z. Ismail, "Implementation of fuzzy time series in forecasting of the nonstationary data," *Int. J. Comput. Intell. Appl.*, vol. 15, no. 2, 2016, Art. no. 1650009.

[25] N. R. Pal and J. C. Bezdek, "On cluster validity for the fuzzy C-means model," *IEEE Trans. Fuzzy Syst.*, vol. 3, no. 3, pp. 370–379, May 1995.

[26] R. J. Hathaway and J. C. Bezdek, "Fuzzy cmeans clustering of in complete dataf," *IEEE Trans. Syst. Man Cybern. B, Cybern.*, vol. 31, no. 5, pp. 735–744, Oct. 2001.

[27] E. Egrioglua, C. H. Aladagb, and U. Yolcuc, "Fuzzy time series forecasting with a novel hybrid approach combining fuzzy c-means and neural networks," *Expert Syst. Appl.*, vol. 40, no. 3, pp. 854–857, 2013.

[28] W. Lu, J. Yang, X. Liu, and W. Pedrycz, "The modeling and prediction of time series based on synergy of high-order fuzzy cognitive map and fuzzy C-means clustering," *Knowl.-Based Syst.*, vol. 70, pp. 242–255, Nov. 2014.

[29] W. Pedrycz and W. Homenda, "Building the fundamentals of granular computing: A principle of justifiable granularity," *Appl. Soft Comput.*, vol. 13, no. 10, pp. 4209–4218, Oct. 2013.

[30] M. Hagiwara, "Extended fuzzy cognitive maps," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Aug. 1992, pp. 795–801.

[31] G. Deschrijver and E. E. Kerre, "On the position of intuitionistic fuzzy set theory in the framework of theories modelling imprecision," *Inf. Sci.*, vol. 177, no. 8, pp. 1860–1866, Apr. 2007.

[32] D. Shan, W. Lu, and J. Yang, "The DataDriven fuzzy cognitive map model and its application to prediction of time series," *Int. J. Innov. Comput. Inf. Control*, vol. 14, no. 5, pp. 1583–1602, 2018.

[33] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[34] A. Domahidi, E. Chu, and S. Boyd, "ECOS: An SOCP solver for embedded systems," in *Proc. Eur. Control Conf. (ECC)*, Jul. 2013, pp. 3071–3076.

[35] G. A. Banini and R. A. Bearman, "Application of fuzzy cognitive maps to factors affecting slurry rheology," *Int. J. Mineral Process.*, vol. 52, no. 4, pp. 233–244, Feb. 1998.

[36] G. Feng, W. Lu, and J. Yang, "The modeling of time series based on least square fuzzy cognitive map," *Algorithms*, vol. 14, no. 3, p. 69, Feb. 2021.

[37] P. Singh, "A brief review of modeling approaches based on fuzzy time series," *Int. J. Mach. Learn. Cybern.*, vol. 8, no. 2, pp. 397–420, Apr. 2017.

[38] M. Bose and K. Mali, "Designing fuzzy time series forecasting models: A survey," *Int. J. Approx. Reasoning*, vol. 111, pp. 78–99, Aug. 2019.

[39] G. Felix, G. Nápoles, R. Falcon, W. Froelich, K. Vanhoof, and R. Bello, "A review on methods and software for fuzzy cognitive maps," *Artif. Intell. Rev.*, vol. 52, no. 3, pp. 1707–1737, Oct. 2019.

[40] G. Feng, L. Zhang, J. Yang, and W. Lu, "Long-term prediction of time series using fuzzy cognitive maps," *Eng. Appl. Artif. Intell.*, vol. 102, Jun. 2021, Art. no. 104274.

[41] O. Orang, P. C. de Lima e Silva, and F. G. Guimarães, "Time series forecasting using fuzzy cognitive maps: A survey," *Artif. Intell. Rev.*, vol. 56, no. 8, pp. 7733–7794, Aug. 2023.

[42] G. Feng, W. Lu, and J. Yang, "Time series modeling with fuzzy cognitive maps based on partitioning strategies," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, Jul. 2021, pp. 1–6.

[43] W. Pedrycz, "Why triangular membership functions," *Fuzzy Sets Syst.*, vol. 64, pp. 21–30, Sep. 1994.

[44] Y. Sheng, H. Liu, and J. Li, "Bearing performance degradation assessment and remaining useful life prediction based on data-driven and physical model," *Meas. Sci. Technol.*, vol. 34, no. 5, May 2023, Art. no. 055002.

[45] Q. Li and Z. Ren, "A new fractional-order augmented quaternion-valued approach for degradation prognostics of bearings using generalized hamilton-real calculus," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.

[46] M. Kolahdoozi, A. Amirkhani, M. H. Shojaeefard, and A. Abraham, "A novel quantum inspired algorithm for sparse fuzzy cognitive maps learning," *Int. J. Speech Technol.*, vol. 49, no. 10, pp. 3652–3667, Oct. 2019.

[47] A. Amirkhani, M. Shirzadeh, M. H. Shojaeefard, and A. Abraham, "Controlling wheeled mobile robot considering the effects of uncertainty with neuro-fuzzy cognitive map," *ISA Trans.*, vol. 100, pp. 454–468, May 2020.
[48] X. Xie, M. Huang, Y. Liu, and Q. An, "Intelligent tool-wear prediction based on informer encoder and bi-directional long short-term memory," *Machines*, vol. 11, no. 1, p. 94, Jan. 2023.
[49] W. Xinxin, S. Xiaopan, A. Xueyi, and L. Shijia, "Short-term wind speed forecasting based on a hybrid model of ICEEMDAN, MFE, LSTM and informer," *PLoS ONE*, vol. 18, no. 9, Sep. 2023, Art. no. e0289161.
[50] M. Gong, Y. Zhao, J. Sun, C. Han, G. Sun, and B. Yan, "Load forecasting of district heating system based on informer," *Energy*, vol. 253, Aug. 2022, Art. no. 124179.
[51] X. Liu, Y. Zhang, J. Wang, H. Huang, and H. Yin, "Multi-source and multivariate ozone prediction based on fuzzy cognitive maps and evidential reasoning theory," *Appl. Soft Comput.*, vol. 119, Apr. 2022, Art. no. 108600.
[52] H. A. Mohammadi, S. Ghofrani, and A. Niksereshtt, "Using empirical wavelet transform and high-order fuzzy cognitive maps for time series forecasting," *Appl. Soft Comput.*, vol. 135, Mar. 2023, Art. no. 109990.
[53] A. A. Tatarkanov, I. A. Alexandrov, L. M. Chervjakov, and T. V. Karlova, "A fuzzy approach to the synthesis of cognitive maps for modeling decision making in complex systems," *Emerg. Sci. J.*, vol. 6, no. 2, pp. 368–381, Mar. 2022.
[54] A. Amirkhani, M. Shirzadeh, T. Kumbasar, and B. Mashadi, "A framework for designing cognitive trajectory controllers using genetically evolved interval type-2 fuzzy cognitive maps," *Int. J. Intell. Syst.*, vol. 37, no. 1, pp. 305–335, Jan. 2022.
[55] A. Amirkhani, M. Kolahdoozi, C. Wang, and L. Kurgan, "Prediction of DNA-binding residues in local segments of protein sequences with fuzzy cognitive maps," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 17, no. 4, pp. 1372–1382, Sep. 2020.

DAN SHAN received the M.S. degree in control theory and control engineering from the Dalian University of Technology, Dalian, China, in 2004. She joined the Dalian Neusoft University of Information, in 2009. She is currently an Associate Professor with the Department of Electronic Engineering. Her current research interests include computational intelligence, fuzzy modeling and granular computing, knowledge discovery and data mining, fuzzy intelligent systems, and low-cost integrated automation systems.

LI WANG received the M.S. degree in microelectronics and solid state electronics from Liaoning University, Shenyang, China, in 2005. She joined the Dalian Neusoft University of Information, in 2017. She is currently a Lecturer with the Department of Electronic Engineering. Her current research interests include computational intelligence, the Internet of Things, and intelligent sensors.

WEI LU (Member, IEEE) received the M.S. and Ph.D. degrees in control theory and control engineering from the Dalian University of Technology, Dalian, China, in 2004 and 2015, respectively. He joined the Dalian University of Technology, in 2004. He is currently a Professor with the School of Control Science and Engineering, Dalian University of Technology. His current research interests include computational intelligence, fuzzy modeling and granular computing, knowledge discovery and data mining, fuzzy intelligent systems, and low-cost integrated automation systems. He also serves as a Frequent Reviewer for many international journals, including IEEE TRANSACTIONS ON FUZZY SYSTEMS, IEEE TRANSACTIONS ON CYBERNETICS, *Knowledge-Based Systems*, *Applied Soft Computing*, *Expert Systems With Applications*, *International Journal of Approximate Reasoning*, *International Journal of Granular Computing* (Springer), as well as international conferences.

JUN CHEN received the Ph.D. degree in materials science from the Dalian University of Technology, Dalian, China, in 2009. He joined the Dalian University of Technology, in 1990. He is currently an Associate Professor with the School of Materials, Dalian University of Technology. His current research interests include evaluation of degradation of mechanical properties of materials, early detection of minor defects in materials, nonlinear ultrasonic testing systems, and bonding structure defect detection.